(72) Inventors:
• Cappelletti, Paolo
20030 Seveso (IT)
• Ricco', Bruno
40126 Bologna (IT)
• Esseni, David
40050 Monterenzio (IT)

(74) Representative:
Cerbaro, Elena et al
STUDIO TORTA S.r.l.,
Via Viotti, 9
10121 Torino (IT)

(54) **Controlled hot-electron writing method for non-volatile memory cells**

(57)   In order to optimise writing of the cell, the latter is written in a condition of equilibrium between the injection current (I_g) and the displacement current ($C_{pp}V_{sl}$). In this way, during writing, the voltage of the floating gate region (V_{fl}) remains constant, as does the drain current and the rise in the threshold voltage. In particular, both for programming and for soft-writing after erasure, the substrate of the cell is biased at a negative voltage (V_{sb}) with respect to the source region, and the control gate region of the cell receives a ramp voltage (V_{cg}) with a selected predetermined inclination (V_{sl}) satisfying an equilibrium condition ($V_{sl} < I_{g,sat}/C_{pp}$).
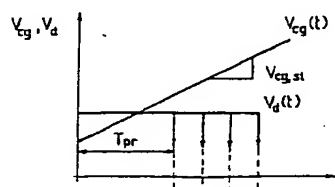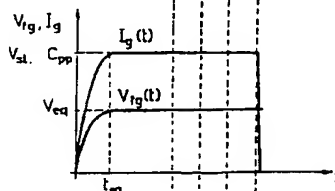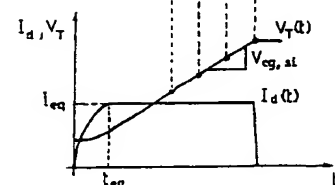
Fig. 3a

Fig. 3b

Fig. 3c

EP 0 908 895 A1

Description

[0001] The present invention relates to a controlled hot-electron writing method for non-volatile memory cells. In particular, for the purposes of the present patent, the term "writing" indicates programming of the cell from an erased condition to a condition for storing a desired threshold value, and the operation of "soft-writing" correcting the threshold voltage of overwritten cells (which involves the partial re-writing of the cells so as to set them all to the same erase threshold voltage).

[0002] As is known, non-volatile memories are becoming increasingly important in modern-day microelectronics, both as separate components and as components forming part of more complex devices; in this context, flash memories play a leading role, which is expected to increase significantly in the future.

[0003] Flash memories (as well as EPROMs) use the technique of hot electron injection for programming of the cells. As is known, programming presents two main problems which are particularly critical in the case of multilevel programming used to store more than one bit for each cell: a) accurate control of the programmed threshold voltages; b) control and limitation of the drain current flowing in the cell during programming.

[0004] The second problem, in particular, directly affects the possibility of operating in parallel on a large number of cells, which is of crucial importance for the operation of "soft-writing" erased flash cells. In general, however, this control function is important for achieving an increase in the performance during the storage of information.

[0005] In general, limitation of the cell current represents only one aspect of the more complex problem of controlling the current since, in order to optimise writing, the drain current should theoretically remain constant during the entire operation, so as to avoid high initial current peaks, followed by lower currents, which result in inefficient programming.

[0006] On the other hand, the problem of accurately controlling the programmed threshold voltage is important for digital memories and is absolutely crucial for multilevel storage, in view of the limited margins which separate the various levels within the available threshold window.

[0007] At present, in order to control the threshold after a programming stage, a verify operation consisting in reading the programmed cell is carried out. This method, however, involves a long and complicated procedure as well as a considerable use of space; consequently, the possibility of controlling in an accurate and reliable manner the threshold voltage of the cell during programming would be highly desirable.

[0008] The conventional writing procedure used for the present generation of flash memories (and EPROMs) uses rectangular pulses for the control gate and drain voltages (V_{cg} and V_d, respectively); consequently, programming is characterized by high drain currents, in particular at the start of the programming pulse, when the overdrive voltage (i.e. the difference between the control gate voltage and the threshold voltage) is high.

[0009] Some solutions have therefore been proposed in order to limit the drain current and/or control the overdrive current, but none of them achieves the ideal result of programming the cell with a constant current.

[0010] The object of the invention is therefore to provide an improved writing method able to operate with a constant drain current, at least for a substantial part of the programming time.

[0011] According to the present invention, a controlled hot-electron writing method for non-volatile memory cells is provided, as defined in Claim 1.

[0012] According to the detailed description given below, in order to optimise writing of the cell, the latter is written in an equilibrium condition, with a constant floating gate region voltage and currents. In particular, both for programming and soft-writing after erasing, the substrate region of the cell is biased at a negative voltage with respect to the source region, and the control gate region of the cell receives a ramp voltage having a slope selected so that it is possible to achieve the condition of an equilibrium between the injection current flowing towards the floating gate region and the displacement current associated with the equivalent capacitor arranged between the floating gate and control gate regions.

[0013] For understanding the invention, a preferred embodiment thereof is now described, purely by way of a non-limiting example, with reference to the accompanying drawings, in which:

- Figure 1 shows a cross-section through a conventional flash cell;
- Figures 2a and 2b show the relation between two electrical parameters of the flash cell of Fig. 1 in two different biasing conditions according to the known method;
- Figures 3a, 3b and 3c show the plot of three electrical parameters of cell with the programming method according to the invention;
- Figure 4 shows the plot of other physical parameters of the flash cell of Fig. 1 with the programming method according to the invention;
- Figures 5a and 5b show the plot of other parameters in a first implementation of the present method;
- Figure 6 relates to other parameters with the first implementation of the present method;
- Figures 7a and 7b show the plot of the parameters of Figs. 5a and 5b in a second implementation of the present method;

- Figure 8 shows the comparative plot of an electrical parameter in the case of programming and soft-writing.

[0014] In order to understand the physical phenomena underlying the present invention, reference should be made initially to Figure 1 showing a conventional flash cell. The flash cell 1 comprises a source region 2 and drain region 3 of the N^+ type embedded in a substrate 4 of the P type having a surface 5 and defining, in the surface portion arranged between the source region 2 and the drain region 3, a channel region 6. A floating gate region 9 extends over the substrate 4, above the channel region 6, and is separated electrically from the latter by a tunnel oxide layer 10. In a known manner, the floating gate region 9 is isolated electrically from all the other conducting regions of the cell. A control gate region 11 extends above the floating gate region 9, isolated electrically from the latter by a dielectric layer 12. The source region 2, drain region 3, control gate region 11 and the substrate 4 are biased through own contact regions, schematically indicated by electrodes 15, 16, 17 and 18.

[0015] On the basis of the conventional capacitive model of the flash cell 1, when the cell is biased so as to be on and the source region 2 is grounded, the coupling between cell regions 2, 3, 9 and 11 is described by equivalent capacitors, shown in Fig. 1. In particular, C_{fs} represents the coupling between the floating gate region 9 and the source region 2; C_{fb} represents the coupling between the floating gate region 9 and the substrate 4 (channel region 6); C_{fd} represents the coupling between the floating gate region 9 and the drain region 3; and C_{pp} represents the coupling between the floating gate region 9 and the control gate region 11.

[0016] In the cell 1, the voltage of the floating gate region V_{fg} is a univocal function of the threshold voltage V_T (voltage of the control gate region at which the cell is activated), of the control gate voltage V_{cg} applied to the control gate region 11 and of the drain voltage V_d applied to the drain region 3 and may be described by the following relation:

$$V_{fg} = V_{T\text{-}fg} + \alpha_{fc}(V_{cg} - V_T) + \alpha_{fd}(V_d - V_{dT}) \qquad (1)$$

in which α_{fc} and α_{fd} are the coupling coefficients, $V_{V\text{-}fg}$ is the floating gate threshold voltage (voltage of the floating gate region at which the cell is turned on), and V_{dT} is the drain region voltage used for measuring V_T.

[0017] The threshold voltage V_T obviously depends on the charge (Q_{fg}) stored in the floating gate region 9 and is typically expressed by:

$$V_T = \frac{V_{T\text{-}fg}}{\alpha_{fc}} - \frac{Q_{fg}}{C_{pp}} - \frac{\alpha_{fd}}{\alpha_{fc}} \cdot V_{dT} \qquad (2)$$

[0018] By substituting (2) in (1), the following well known equivalent expression for the voltage V_{fg} is obtained:

$$V_{fg} = \alpha_{fc} V_{cg} + \alpha_{fd} V_d + \frac{Q_{fg}}{C_T} \qquad (3)$$

in which the identity $\alpha_{fc}/C_{pp} = 1/C_T$ has been used, and $C_T = C_{pp} + C_{fs} + C_{fd} + C_{fb}$.

[0019] During the programming stage, V_{cg}, V_d and Q_{fg} generally vary with time. In particular, the time derivative of $V_T(t)$ may be obtained from (2) and is equivalent to:

$$\dot{V}_T = \left[\frac{dV_T}{dt}\right] = -\frac{\dot{Q}_{fg}}{C_{pp}} = \frac{I_g}{C_{pp}} \qquad (4)$$

in which I_g indicates the injection current (Fig. 1) and by definition $I_g = -\dot{Q}_{fg} = -[dQ_{fg}/dt]$.

[0020] Moreover, deriving over time (3) gives:

$$\dot{V}_{fg} = \left[\frac{dV_{fg}}{dt}\right] = \alpha_{fg}\dot{V}_{cg} + \alpha_{fd}\dot{V}_d - \frac{I_g}{C_T} = \frac{1}{C_T} \cdot [C_{pp}\dot{V}_{cg} + C_{fd}\dot{V}_d - I_g] \qquad (5)$$

[0021] In the case of conventional programming, in which V_{cg} and V_d have a rectangular form, (5) is reduced to:

$$\dot{V}_{fg} = -\frac{I_g}{C_T} \qquad (5')$$

EP 0 908 895 A1

and the injection current I_g has the qualitative behavior shown in Fig. 2a. During programming, V_{fg} is a decreasing monotone function (as shown schematically by the direction of the arrow along the abscissa) and therefore the drain current I_d of the cell and the programming velocity $\dot{V}_T$ also decrease monotonically.

[0022] On the other hand, according to the present method, the control gate voltage is made to increase linearly with a predetermined slope $\dot{V}_{cg} = V_{sl}$ and at the same time a rectangular drain voltage is applied; in this condition the floating gate voltage V_{fg} increases until it reaches an equilibrium condition and further programming of the cell is performed at a virtually constant floating gate voltage ($V_{fg} = V_{eq}$); moreover, the drain current I_d increases monotonically with V_{fg} up to the equilibrium value (as shown in Fig. 2b) and therefore remains constant.

[0023] In fact, by biasing the cell as described, (5) becomes:

$$\dot{V}_{fg} = \frac{1}{C_T} \cdot [C_{pp}V_{sl} - I_g(V_{fg})] \qquad (6)$$

in which the drain voltage is considered constant.

[0024] The essential evolution of the programming stage may be described as follows. At the start of the ramp of the control gate voltage, when V_{fg} is relatively small, the injection of hot electrons is very low; therefore the second term of (6) may be disregarded and V_{fg} increases with V_{cg} owing to the effect of the first term of (6). Assuming that the injection current I_g is a greatly increasing function of V_{fg} (as discussed in detail below), $\dot{V}_{fg}$ decreases gradually until it becomes zero when I_g is equivalent to $[C_{pp}V_{sl}]$.

[0025] At this point, the floating gate voltage V_{fl} has reached its equilibrium value V_{eq} corresponding to the equilibrium between the injection current I_g and the displacement current $[C_{pp}V_{sl}]$, therefore the floating gate voltage remains constant even in the presence of further variations in the control gate voltage V_{cg}.

[0026] Qualitative waveforms for V_{cg} and V_d according to the invention are shown in Fig. 3a; Fig. 3b shows the behavior of the voltage V_{fg} and the injection current I_g. In these figures, T_{pr} represents the programming time, while t_{eq} indicates the time necessary to reach the equilibrium condition $I_g(V_{fg}) = C_{pp}V_{sl}$. Figure 3c shows the behavior of the threshold voltage V_T and the drain current I_d; as shown, after reaching the equilibrium, since V_{fg} and V_d are constant, the drain voltage also becomes constant and hence $I_d = I_{eq}$. Moreover, as can be seen clearly from (4), assuming $I_g = C_{pp}V_{sl}$, $\dot{V}_T = V_{sl}$ is obtained, and hence the threshold voltage V_T, in the equilibrium condition, follows exactly the trend of the control gate voltage, imposed externally.

[0027] From the above it is clear that two essential conditions exist for achieving the equilibrium condition and hence programming of the cell with a constant drain current, namely that the injection current I_g should be a monotonically increasing function of V_{fl}, preferably greatly increasing, and that the slope of the control gate voltage V_{sl} should be such that it is possible to reach the condition of equilibrium between the injection current and the displacement current, namely that the injection current I_g, during its increase, is able to become equal to the displacement current $C_{pp}V_{sl}$.

[0028] The first condition (injection current increasing, preferably rapidly, with the floating gate voltage) is ensured by the type of programming (hot electron injection) and by biasing the substrate with a negative voltage, as discussed below in detail; the second condition requires that the slope of the ramp of the control gate voltage should not be too high. In particular, since the injection current I_g tends to become saturated when the floating gate voltage V_{fl} nears the drain voltage V_d (see Fig. 4, showing the experimental results), this condition requires that the slope V_{sl} should be less than a maximum value namely

$$V_{sl} < I_{g,sat}/C_{pp}, \qquad (7)$$

with $I_{g,sat}$ the saturation value.

[0029] Biasing the substrate (body) of the cell at a negative voltage with respect to the source region 2 has the purpose, on the one hand, of reducing and improving control of the value of the drain current used during writing and, on the other hand, of increasing the injection current, so that it is a greatly increasing function of the floating gate voltage. In this way, the first condition described above for achieving the equilibrium situation is satisfied and programming is performed with an injection characteristic $I_g(V_{fg})$ suitable for rapidly reaching the equilibrium, ensuring that it is reached at an early stage of writing and in any case before the end thereof. The application of a voltage ramp to the control gate region and a negative substrate voltage therefore operates in a synergic manner so as to optimise the programming operation, as is clear from the results of the experiments described below.

[0030] In particular, the present method envisages biasing the substrate with the source-body voltage V_{sb} ranging between 0.5 and 4 V, depending on the type of device, the operating voltages thereof and its value may be optimised on the basis of the results of Fig. 4.

[0031] The application of a negative substrate voltage (source-body voltage $V_{sb} > 0$) causes an increase in the injection current, as shown in Fig. 4, which shows the characteristics of the injection current I_g and drain current I_d for a cell

4

of 0.6 μm for V_{sb} = 0 and V_{sb} = 3 V. The figure also shows the equilibrium values of V_{fg} corresponding to slopes V_{sl} = 1 mV/μs and 100 mV/μs. In particular, from the figure it will be noted that it is possible to highlight a range of floating gate voltages V_{fg} in which there is an improved behaviour of the injection currents and drain current and in which in particular the injection current is greatly increasing with V_{fg}, so as to ensure the conditions for reaching the equilibrium.

[0032] The results of the experiments carried out have confirmed all the qualitative characteristics described above as regards both programming and soft-writing of the cell.

[0033] In particular, when writing is used for multilevel programming it is necessary to obtain different thresholds with a same injection mechanism, which may be obtained either using pulses with a different amplitude and the same duration or pulses with the same amplitude and different duration. In the case of the ramp form used here, it is possible to obtain different threshold voltages by simply varying the width of the drain pulse, since, when V_d falls low again, the threshold voltage no longer rises and the programming operation is interrupted. Fig. 3b shows how it is possible to obtain an arbitrary threshold voltage using a same waveform of the control gate voltage by adjusting the duration of the pulse T_{pr}.

[0034] Results obtained during the programming of cells with a control gate voltage V_{cg} having a positive initial value, slope V_{sl} of 100 mV/μs (corresponding to an equilibrium drain current I_{eq} of 50μA and a floating gate voltage V_{fg} ≅ 3.1 V) are shown in Figs. 5a and 5b, in which the different characteristics correspond to different values of the initial threshold voltage. It is significant to note that the use of the ramp voltage eliminates the presence of current peaks (or "spikes") which are typical of the conventional method with rectangular control gate voltage and without biasing the substrate. In this connection reference should be made also to Fig. 6 in which the waveforms have been adjusted so that the thresholds and the programming times are approximately the same. As is known, the new method of programming allows a maximum drain current I_d five times smaller than that of the conventional method to be obtained.

[0035] When writing is used to recover overwritten cells it is of primary importance to reduce the drain current I_d flowing in each cell and control precisely the threshold voltage $V_{T,min}$ towards which the cells must converge. In this case, during experimental tests carried out, a slope of V_{sl} = 1 mV/μs was used, corresponding to a soft-writing programming time of 3 ms, a current of 1.5 μA and an equilibrium voltage V_{eq} of 2.1 V. The experimental results obtained are shown in Figs. 7a and 7b. In general, the soft-writing programming time and current consumption may be adjusted over a wide range of values and it is possible to estimate with great accuracy the drain current for a given value of the slope using the data of Fig. 4.

[0036] As will be noted from Fig. 7b, after an initial transient corresponding to t_{eq} of Fig. 3b, the current I_d is virtually independent of the initial threshold voltage which influences only the moment when writing starts. Moreover, when a cell has a threshold voltage greater than $V_{T,min}$, its threshold is not altered by the writing procedure.

[0037] Moreover, since the drain current I_d of each cell is virtually constant over time and independent of the initial threshold value and since the cells with an initial voltage greater than the desired value are not activated, the current used by a group of cells may be easily estimated knowing the initial distribution of the erase thresholds.

[0038] Fig. 8, finally, shows the plot of the equilibrium drain current I_{eq} versus V_{sb} in the case of programming with T_{pr} = 45 μs (curve a) and soft-writing with T_{pr} = 3ms (curve b), for a drain voltage V_d = 4 V. As will be noted, for a given drain voltage, the use of a negative substrate voltage reduces monotonically the current necessary for programming or rewriting the cell in a given time period. As may be noted, the reduction in current has a similar plot in both cases, but the effect is particularly evident in the case of soft-writing, where a voltage V_{sb} = 3 V provides a reduction in current about forty times greater than in the case of V_{sb} = 0 V.

[0039] The ramp form may be obtained by means of an analog block (typically an integrator) or, for reasons associated with the practical implementation, the ramp may be emulated by a small-step waveform. The step approximation results in a current which has a ripple evolving about a mean value fixed by the (mean) slope of the stepped waveform itself. It is therefore clear that the application of a stepped waveform involves a deviation from the ideal behaviour and therefore imposes an upper limit on the duration and the amplitude of the steps, a limit which is fixed by the need to limit the band of values inside which the drain current must in any case be maintained.

[0040] In particular, if the transconductance of the cell in the equilibrium condition (to which the equilibrium drain current I_{eq} corresponds) is called $G_{m,eq}$ = (dI_{eq}/dV_{cg}) and if ΔV_{cg} indicates the step amplitude in the ramp emulation, the current variation ΔI_{eq} can be expressed as ΔI_{eq} = $G_{m,eq} \cdot \Delta V_{cg}$. If the current percentage variation is required to be less than a predetermined value V% = $(\Delta I_{eq}/I_{eq})$, it is therefore necessary to ensure that:

$$\Delta V_{cg} < I \ (I_{eq}/G_{m,eq}) \cdot V\%$$

[0041] Once the values of ΔV_{cg} have been determined, all the parameters of stepped waveform are fixed since, if T_{pr} indicates the programming time and V_{cgSW} the overall voltage swing at the control gate region, the number of steps must be equal to N = $V_{cgSW}/\Delta V_{cg}$ and the duration of the individual step is equal to T_{pr}/N.

[0042] This criterion provides in practice steps with a duration ranging between 1 μs and 10 ms depending on whether writing relates to programming of the cell or soft-writing.

[0043]   The advantages of the described method are the following. Firstly, writing of the cells in a condition of equilibrium between the injection current and the displacement current (and, hence, as described, with constant floating gate voltage, drain current and threshold voltage increase), produces continuous control of the programming parameters and hence optimum programming of the cells. Since, in the equilibrium condition, the drain current is controlled and constant, additional current-limiting circuits are not necessary and hence the associated circuitry is simplified. Accurate control of the threshold voltage and drain current by the slope of the voltage ramp applied to the control gate region allows both rapid and accurate writing of multiple threshold voltages during the programming of multilevel memories and slow writing with low-consumption and high parallelism of a plurality of cells for soft-writing. Moreover, the biased substrate ensures that the cell may reach the equilibrium condition rapidly and reliably, acting in synergism with application of the ramp gate voltage.

[0044]   The method has a high degree of flexibility, may be used in a dynamic and adaptable manner during the test and during operation of the cells, by choosing the slope of the ramp so as to optimise the programming time and current consumption within a wide range of values, and may therefore be used both for digital and multilevel programming and for the soft-writing of cells after erasure.

[0045]   Finally it is obvious that numerous modifications and variations may be made to the method described and

[0046]   Finally it is obvious that numerous modifications and variations may be made to the method described and illustrated herein, all of which falling within the scope of the inventive idea, as defined in the accompanying claims.

Claims

1.   A controlled hot-electron writing method for non-volatile memory cells (1) comprising a source region (2), a drain region (3) separated from said source region by a channel region (6) formed in a semiconductor material body(4), a floating gate region (9) extending over at least part of said channel region and set to a floating gate voltage (V_{fl}) and a control gate region (11) extending over at least part of said floating gate region, said method comprising the step of applying biasing voltages to said drain region (3) and control gate region (11), characterized in that said floating gate voltage (V_{fl}) is maintained substantially constant during at least part of the programming time.

2.   A method according to Claim 1, characterized in that said step of maintaining comprises the steps of applying to said control gate region (11) a ramp voltage (V_{cg}) with a predetermined mean slope and applying to said body of semiconductor material (4) a negative voltage (V_{sb}) with respect to said source region (2), said predetermined mean slope (V_{cg}) being less than the ratio of the saturation injection current ($I_{g,sat}$) flowing from said channel region (6) to said floating gate region (9) in the saturation condition and the coupling capacitance (C_{pp}) between said floating gate region (9) and control gate region (11).

3.   A method according to Claim 2, characterized in that said ramp voltage (V_{cg}) is increased linearly in a continuous manner.

4.   A method according to Claim 2, characterized in that said ramp voltage (V_{cg}) is increased linearly in a discrete manner and forms a plurality of steps.

5.   A method according to Claim 4, characterized in that each step has a duration ranging between 1 µs and 10 µs.

6.   A method according to any one of Claims 2 to 5, characterized in that said negative voltage (V_{sb}) ranges between -0.5 and -4 V.

7.   A method for programming the threshold of non-volatile memory cells according to any one of the preceding claims.

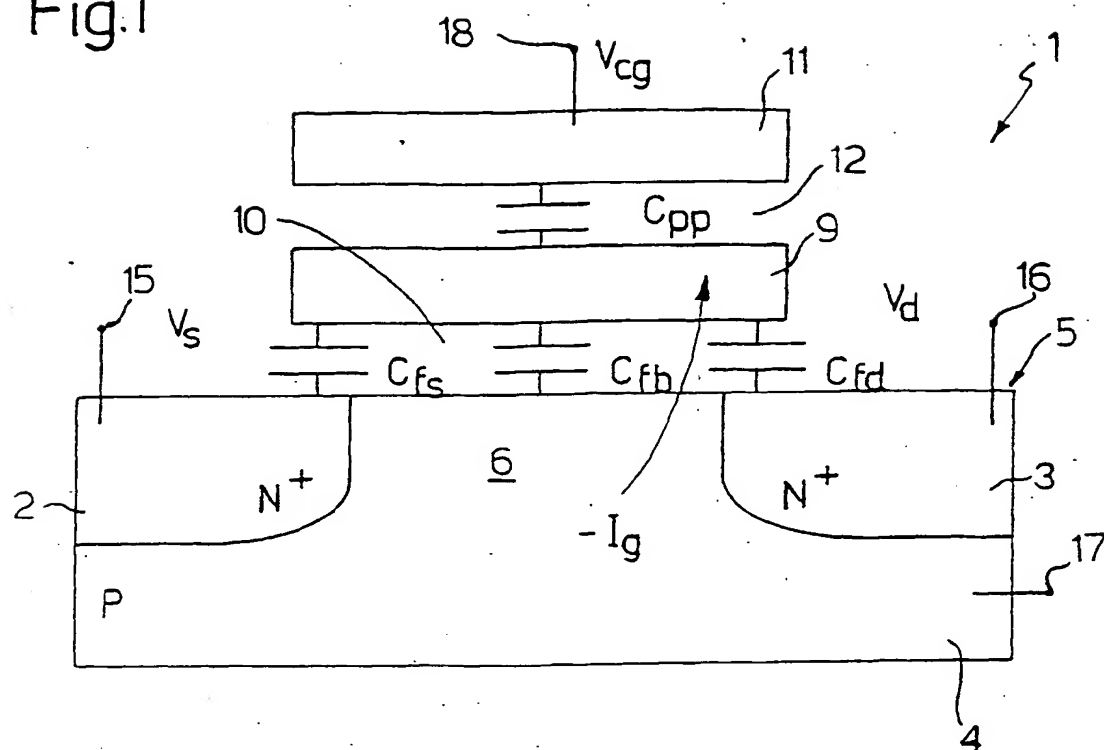8.   A method for soft-writing non-volatile memory cells according to any one of the preceding claims.
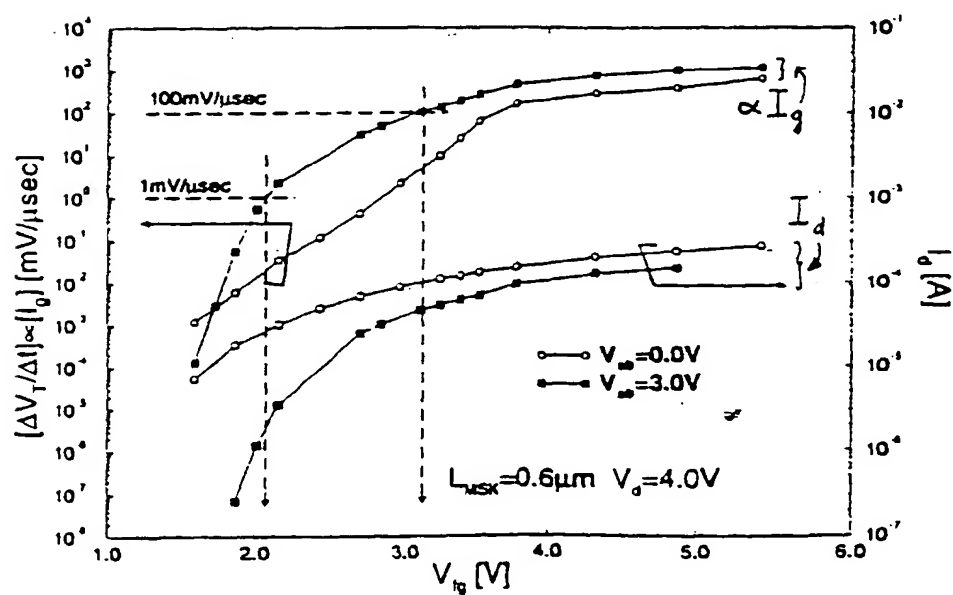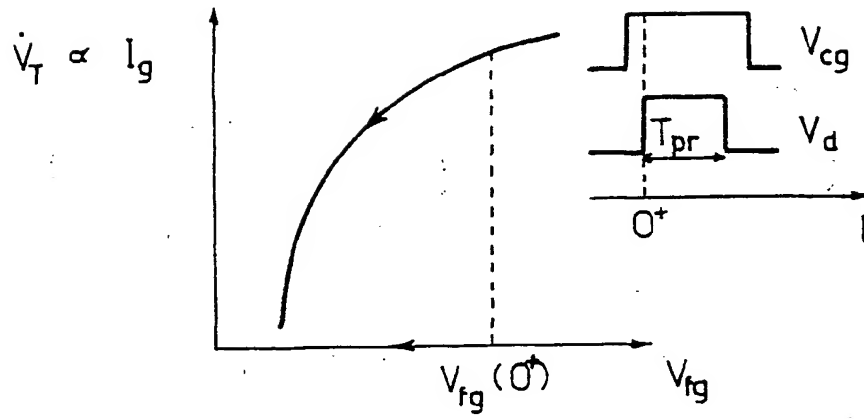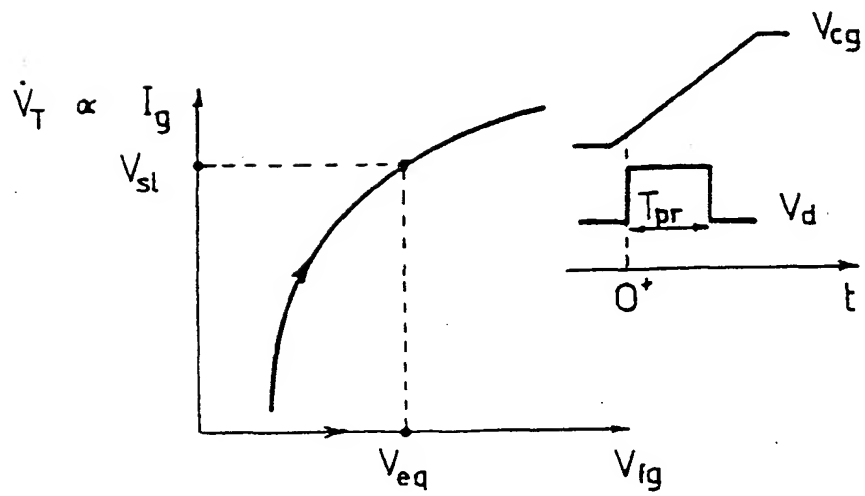
6

Fig.1

18 V_{cg} 11

1

12 C_{pp}

10 9

15 V_s

16 V_d

5

C_{fs} C_{fb} C_{fd}

2 N^+ $\underline{6}$ N^+ 3

$-I_g$

17

P

4

Fig.4

10^4 10^{-1}

10^3

$100mV/\mu sec$ 10^2 $\propto I_g$ 10^{-2}

10^1

$1mV/\mu sec$ 10^0 I_d 10^{-3}

10^{-1}

10^{-2} 10^{-4}

10^{-3}

10^{-4} 10^{-5}

$[\Delta V_T/\Delta t] \propto [I_g]$ [mV/μsec]

I_d [A]

10^{-5}

10^{-6} $V_{sub}=0.0V$

$V_{sub}=3.0V$

10^{-7}

$L_{MSK}=0.6\mu m$ $V_d=4.0V$

10^{-8}

1.0 2.0 3.0 4.0 5.0 6.0

V_{fg} [V]

Fig.2a

Fig.2b
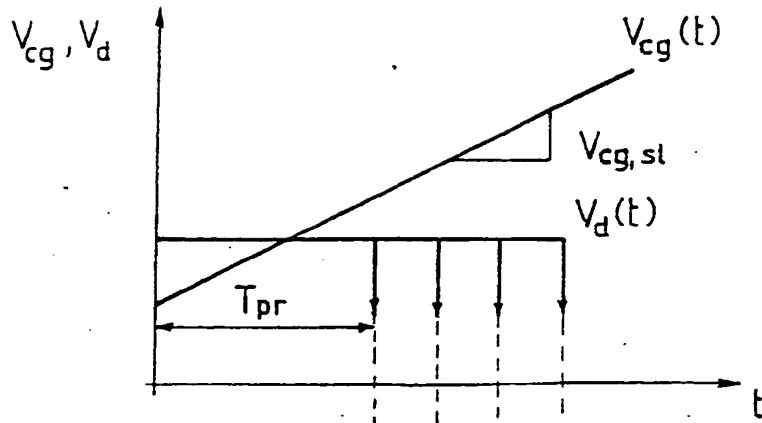
Fig. 3a
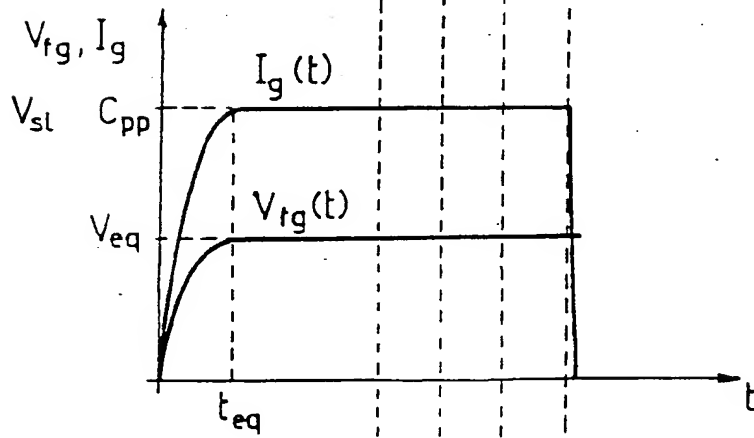
Fig. 3b
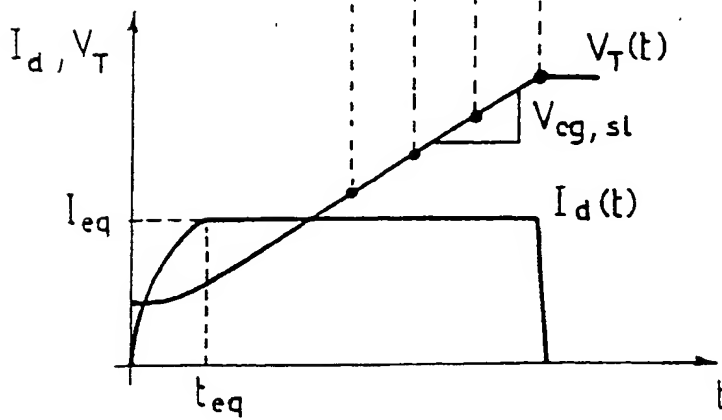
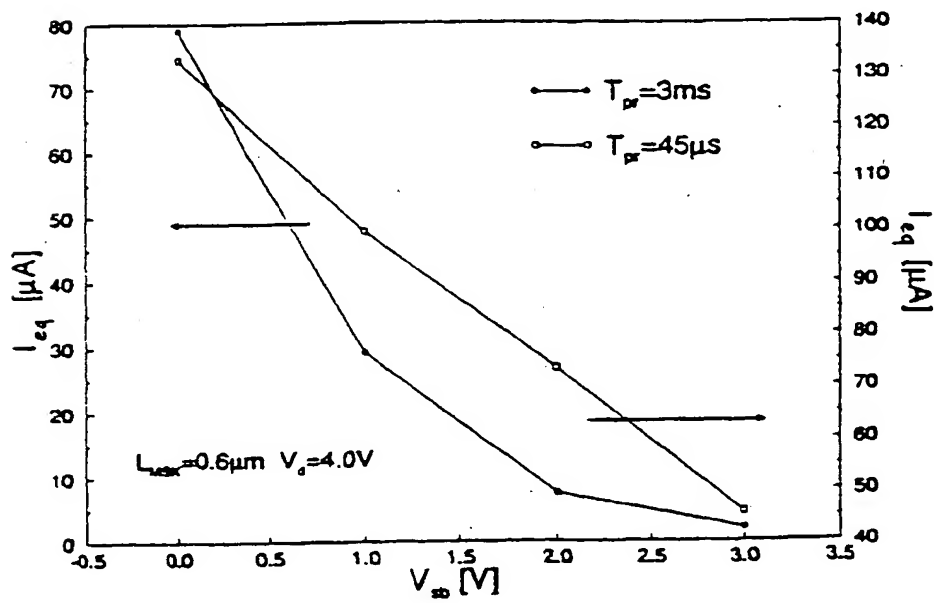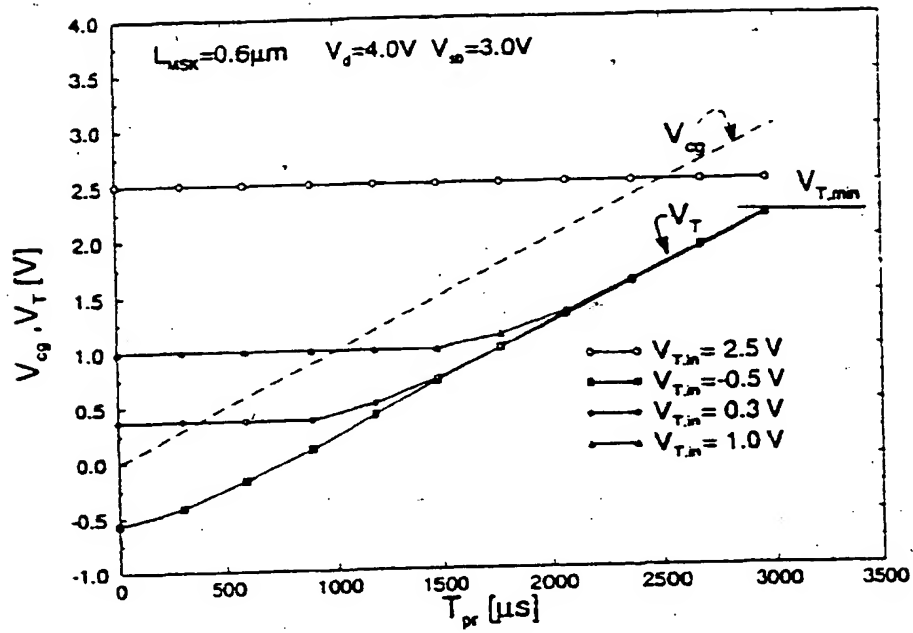Fig. 3c
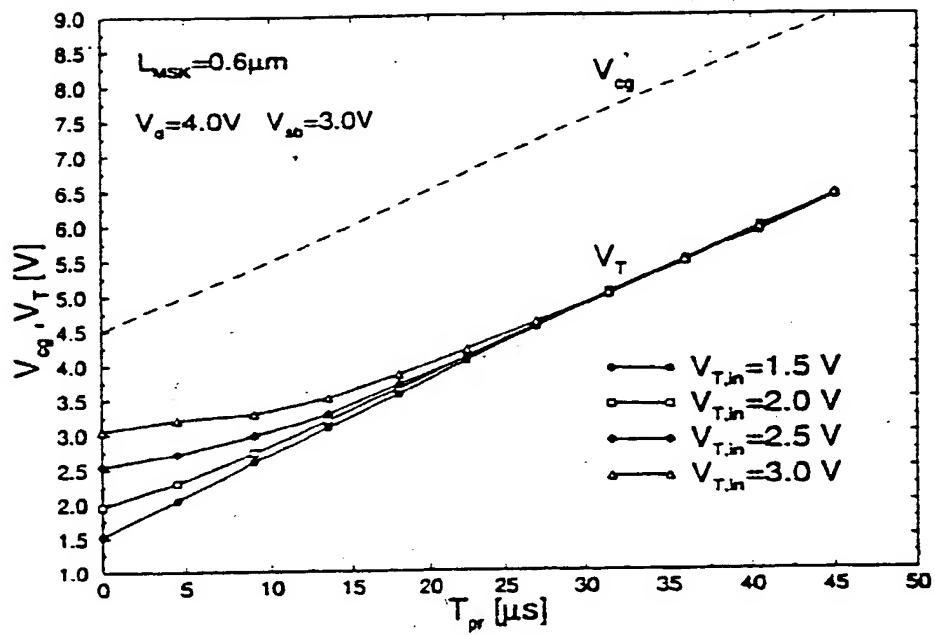
Fig.5a
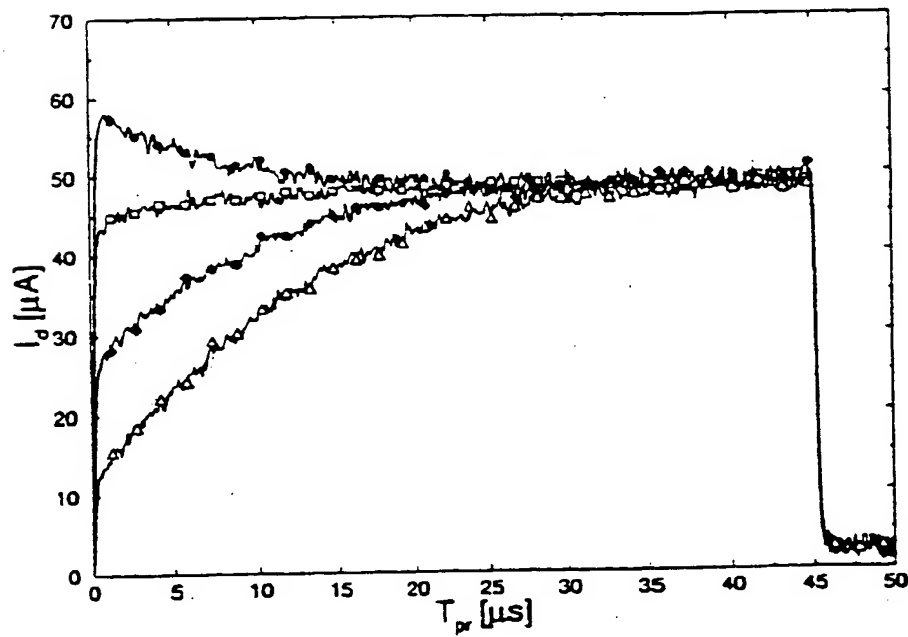
Fig.5b
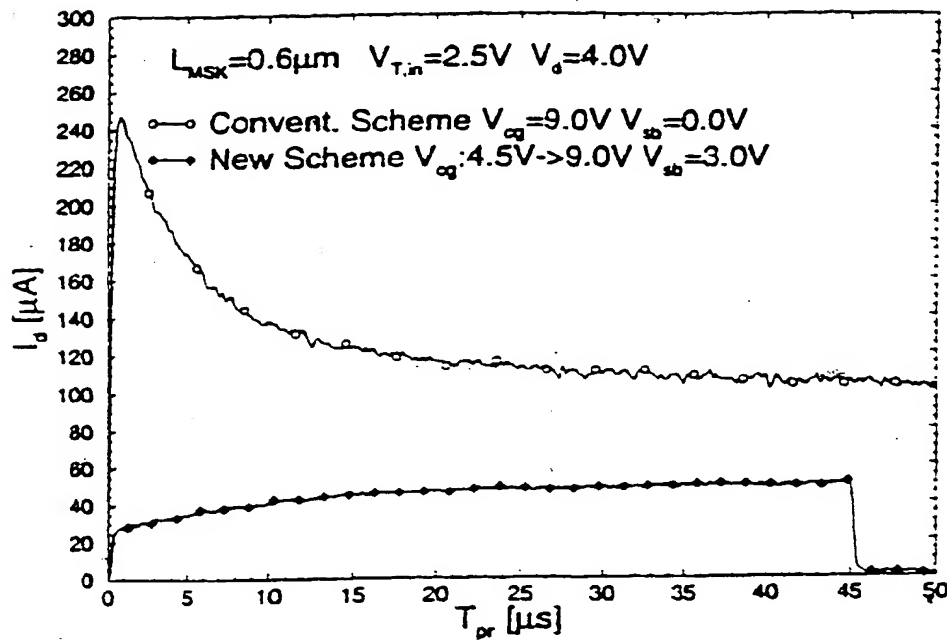
Fig.6
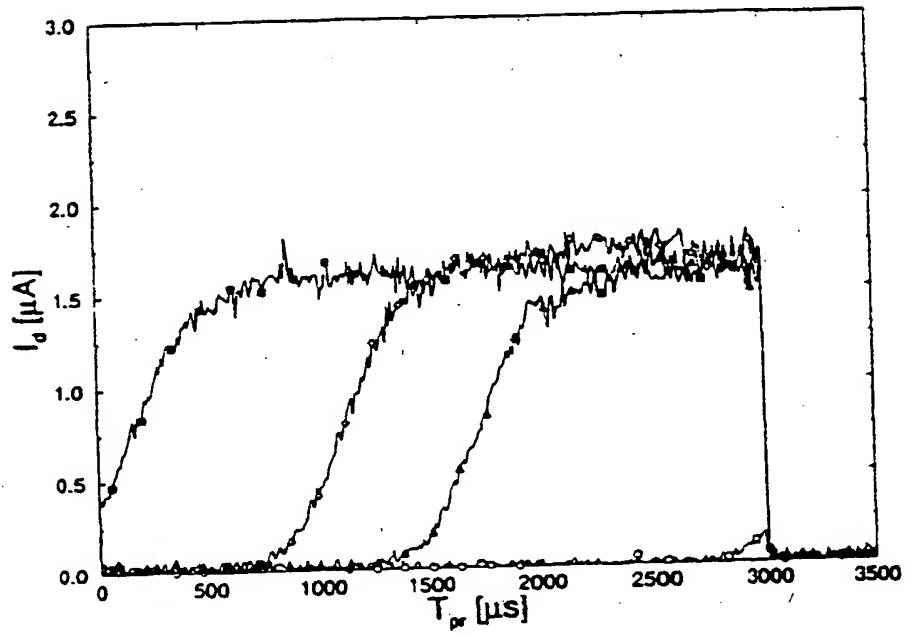

Fig.8

Fig.7a

Fig.7b

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 97 83 0504

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
X	US 5 487 033 A (KEENEY STEPHEN N ET AL) 23 January 1996	1,7,8	G11C16/00 G11C16/06
A	* column 2, line 42 - column 3, line 25; figure 8 * * column 6, line 49 - column 7, line 33 * ---	2-4,6	
X	EP 0 463 331 A (TEXAS INSTRUMENTS INC) 2 January 1992	1,7	
A	* column 1, line 44 - column 2, line 33; figures 1-3 * * column 4, line 12-23; claims 10,17 * ---	3,4	
X	US 5 481 492 A (SHOEMAKER PATRICK A) 2 January 1996 * column 1, line 43 - column 2, line 5 * * column 3, line 46 - column 5, line 20 * -----	1,7	

	TECHNICAL FIELDS SEARCHED (Int.Cl.6)
	G11C

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	11 June 1998	Czarik, D

EPO FORM 1503 03.82 (P04C01)

13

THIS PAGE BLANK (USPTO)